Proc. of the IWAC2022
pp. 170-175, Oct. 2022

Paper ID : T10-1-A

Automatic speech recognition in noise polluted cockpit environments for monitoring the approach briefing in commercial aviation

Sven Bollmann[1)], Jonas Füllgraf[1)], Christian Roxlau[1)], Thomas Feuerle[1)], Peter Hecker[1)], Aravind Krishnan[2)], Simon Ostermann[2)], Dietrich Klakow[2)], Nicolas Großmann[3)], and Stefan Müller-Divéky[4)]

[1)] *Institute of Flight Guidance, Technical University of Braunschweig, Germany*
[2)] *DFKI GmbH, Stuhlsatzenhausweg 3, Saarbrücken, Germany*
[3)] *DFKI GmbH, Trippstadter Straße 122, Kaiserslautern, Germany*
[4)] *Diehl Aerospace GmbH, Germany*
[1)] *email: s.bollmann@tu-braunschweig.de*

The approach briefing is of major importance in commercial aviation. Conducted by the flight crew, it ensures a thorough and mutual understanding of the upcoming descent and approach phase. With regard to a future implementation of reduced crew operations (RCO), an AI-based system is currently being developed that is able to follow the spoken approach briefing, check for its completeness and inform the pilot about possibly missing items. This paper describes the language processing part of the overall system. A commercially available automatic speech recognition system is trained on aviation specific vocabulary and strategies for dealing with cockpit noise are discussed. Steps towards a possible certification of the system according to the European Union Aviation Safety Agency (EASA) Artificial Intelligence Roadmap are outlined.

Key Words : Automatic speech recognition, Approach briefing, Noise handling, Reduced crew operations

1. Introduction

Situational awareness is a key requirement in aviation safety. Highly sophisticated avionic systems, standard operating procedures as well as clear communication and a thoroughly trained cockpit crew ensure that the current flight situation as well as its projection into the future are fully understood by the pilots.

A commonly used technique to create mutual situational awareness among the pilots are briefings. They cover the upcoming flight phase like e.g. the departure or the approach and are currently held freely by one of the pilots and monitored by the other crew member. The underlying structure, however, is standardized. This is to make sure that all relevant topics within the briefing are addressed.

From a more general perspective, briefings can be thought of as narrated situational awareness. It is assumed that pilots are aware of the topics they have talked about in their briefings. If a certain topic has not been addressed, however, there might be a risk that the pilot conducting the briefing was actually unaware of this topic. A standardized briefing structure helps to mitigate this risk. In situations with low workload this usually works well. Things may be different in high workload situations like a sudden change of the runway direction. In such cases important briefing topics might be inadvertently omitted, which could cause a safety issue. In this paper the current development of a system is described that is able to monitor the spoken approach briefing of pilots and check for its completeness against a reference briefing structure. When the briefing is completed, the system informs the pilots about potentially missing items that have not been addressed yet. These items can then be rebriefed by the cockpit crew which will improve the necessary situational awareness for the upcoming approach phase.

The paper is organized as follows: In section 2 the basic structure and contents of a typical approach briefing are explained. Section 3 describes the concept of operations for the overall system in which the speech recognition part is embedded. How aviation specific training data for the automatic speech recognition system was obtained is explained in section 4. The following section 5 describes the training of the language model which forms the core of the speech recognition system used in this project. Preliminary results with noise free test data are presented in section 6. Since the speech recognition task will be performed in a noisy cockpit environment, strategies for dealing with noise are elaborated in section 7. A detailed description of how the recognized speech will be processed afterwards is given in section 8. The necessary steps towards a possible certification of such a system by the European Union Aviation Safety Agency are summarized in section 9. An outlook on future briefing techniques is given in section 10.

All efforts that have been made in the scope of this paper are part of the ViCKI project that is funded by the German Federal Ministry for Economic Affairs and Climate Action. The overall project is aimed on the development of a Virtual Crew Assistent (VCA) using artificial intelligence.

2. Structure of the approach briefing

In general, within commercial air transport, the structure and content of the approach briefing depend on the aircraft type and airline specific procedures. In this context most aircraft manufacturers provide a reference briefing structure which is normally used. However, operators are free to customize the reference to cover special needs in terms of their daily operational environment. Against this background and to avoid company- and aircraft type specific briefing structures, the authors agreed to use a generic briefing structure as a reference for this work. During the data collection, participating pilots were asked to use this structure.

The reference briefing is divided into eight different sections. Some of them have mandatory subtopics, others stand on their own and may be covered more freely. The sections are:

Technical Status - Within this first section the technical status of the aircraft is addressed. In particular, malfunctions which lead to operational restrictions for the upcoming flight phase are reviewed, e.g. entries on the Hold Item List (HIL).

NOTAM - In this section all relevant Notices to Air Missions (NOTAMs) for the approach, landing and go-around are addressed. NOTAMs are publications to inform personnel involved in flight operations about short-term changes to any aeronautical service, facility, procedure or hazard with possible influence on the flight execution.

Weather - Here, the weather at and along the way to the destination and the alternate airport is addressed. Subtopics include visibility, clouds, precipitation, winds and special weather phenomena.

Fuel - This section deals with all necessary fuel considerations regarding remaining fuel at the destination, fuel for go-arounds at the destination airport and the minimum fuel for a diversion to the alternate airport.

Descent - This section covers important aspects of the descent phase like e.g. mandatory waypoints along the arrival route, altitude and speed constraints, the surrounding terrain, the respective altitudes to ensure terrain clearance and so on.

Holding - Within this section the probability to fly a holding procedure is assessed. Flight crews expect holdings especially in high traffic situations or when the destination airport is not accessible, e.g. due to bad weather.

Approach - This section addresses the approach within the terminal maneuvering area (TMA) and has several subtopics that are mandatory to discuss. These include the type of approach, the final descent point, the glide path angle, minimum descent altitudes for the approach, the missed approach procedure and alternate considerations.

Landing - Here, all relevant information about the landing itself is addressed. This includes the dimensions and condition of the runway and the expected wind. Furthermore, the landing distance available (LDA) is compared to the required landing distance (RLD). The use of autobrakes and thrust reversers as well as the expected taxi route are also discussed.

3. Concept of operations

When designing a system like the proposed approach briefing assistant, it is important to carefully determine how such a system will be operated by pilots. In the current project phase the system is assumed to be operated in a conventional cockpit with two pilots. Normal operations will not be affected by the system. If the pilots wish that their briefing shall additionally be monitored by the system, they manually invoke it by pushing a dedicated button. Once the briefing of the pilot flying is completed, the pilot monitoring brings up topics that might have remained unclear to him. Only when no further clarifications are needed, the pilots invoke the report page of the briefing assistant to check for items that have possibly been omitted.

A similar system could also be used in reduced crew operations or even single pilot operations. While the basic functionality will remain the same, the consequences of a system failure differ drastically. In this case the (single) pilot has to fully rely on the error-free functioning of the approach briefing assistant, since no other human pilot is involved in the briefing process.

4. Collecting training data for ASR

For the initial training of the language model the acquisition of training data was necessary. To this end, interviews with type-rated A320 pilots were conducted via an online conference tool. At the beginning of each interview session, pilots were introduced to the project goal and the reference briefing structure, as they might use a different structure in their operational business. In addition, general requirements for the desired briefing were explained, including the exclamation of section names, the language (English) and the request to speak briefly, concisely and in their own words.

After this introduction a specific scenario was presented to the pilots. It describes a vectored straight-in instrument approach to Frankfurt Airport during fair weather conditions with a technical fully functional aircraft. The assumed traffic situation was low. An instrument approach chart for the ILS approach runway 25L was provided to the participants along with further details on the current position of the aircraft and its landing performance. After some preparation time, during which the pilots could familiarize themselves with the situation, the approach briefing was carried out and recorded in an audio file. The recorded briefings were later transcribed manually. A total of 14 briefings were collected this way. In order to expand the training data base, the briefings were later modified manually in terms of word choice and phrasing. Eventually, a training data set with 44 different approach briefings could be obtained.

5. Training the language model

Before processing the approach briefings on a textual level, it is necessary to use methods of Automatic Speech Recognition (ASR) in order to transform speech signals into text. We use an out-of-the-box, state of the art solution provided by Cerence, the VoCon ASR system. An essential building block of such ASR systems is a language model.

A language model is in a certain sense a grammar that defines the words that are relevant for a specific application and how those words can be combined into sentences. More precisely, it is a model that defines a probability distribution over words and sequences of words, and can thus specify what the probability of a sentence is. Language models are typically trained on large amounts of application specific training data. Such data doesn't exist in the context of approach briefings. Therefore, the central problem pivots around constructing a domain-adapted language model without compromising plasticity. The major hurdles involved in building language models for approach briefings are the regular use of domain specific keywords/acronyms (which are otherwise not used in conversational English) and the deployment of a slightly distorted version of commonplace grammar. These make it difficult to use most publicly available resources to train language models for briefings.

Language modelling techniques that can work with small training sets are predominantly explored in this context. This involves the construction of two independent models: a *background* model, which can vaguely generalize the domain of application, and an *adaptation* model which is extremely specific to the task at hand. The closest conversational schema that shares traits with approach briefings is Air Traffic Communication (ATC). Therefore, an ATC corpus is compiled using transcripts from ATCOSIM[5)] datasets. The background model is then generated from this ATC corpus. The adaptation model is generated with the 39 briefing samples. Each briefing is considered as a single sequence of tokens when building the adaptation model. For experimentation, we call this the *ViCKI corpus*. The language models used are backoff 3-gram models built using the SRILM[10)] toolbox. Language modelling experiments are carried out with background and adaptation models, acting individually and in combination. Specifically, we experiment with brute combination and linear interpolation. Brute combination involves merging the background and adaptation corpora and training a model on the resulting corpus. In linear interpolation the background and adaptation model are additively combined on a probabilistic level. That is, the probability of the occurrence of a word sequence is determined by a scaled addition of its corresponding probability in the background and adaptation model.

6. Preliminary results

The 44 documents collected come from 13 interviews in total. 13 documents are therefore transcriptions of real briefings and 31 documents are augmented versions of these. The augmented versions are not accompanied by acoustic data and therefore can only be used in Language model evaluations. We keep augmented versions together with the original briefings in both train and test sets to avoid leakage. A train test split ratio of 90:10 is used throughout the experimentation phase.

For the language model, we use *perplexity* as the evaluation metric. Perplexity is mathematically expressed as the normalized inverse probability of a test example. Essentially, it measures how perplexed/surprised a language model is about the existence of a test sentence. The lower the ”surprise”, the better it understands language of similar construct. The language model perplexities using 11-fold cross validation on the entire dataset is shown in Table 1. An interpolation coefficient (λ) of 0.3 is empirically seen to work well for linear interpolation. The vocabulary of the language model spans every word that appears more than once in ATC or ViCKI corpora. Language models are directly plugged into the VoCon system and used for ASR applications. The ASR module is tested and evaluated using three briefings. The metric used to evaluate ASR performance is Word Error Rate (WER). It measures the number of additions and substitutions required to match the predicted text with its corresponding ground truth. As for perplexity, a lower WER indicates better recognition.

The Word Error rates obtained when the language models are plugged into the VoCon system is also shown in Table 1. It can be seen that the interpolated model gives the best Word Error Rate for the test utterance, while the ViCKI Trigram model results in the best perplexity. The model that is trained on ATC only shows very bad performance, which emphasizes the need for in-domain training data, even if at small scale.

7. Dealing with noise

The task to recognize speech in a cockpit environment is quite challenging. Apart from the previously described aviation related terminology, the noise level during flight is significant. Sources of noise include airflow, engine noise, mechanical noise from actuators, levers, buttons and switches as well as communication with air traffic control or any kind of interfering speech from other sources.

In many ASR systems all kinds of background noises are incorporated in the acoustic model. This is also the case for the VoCon ASR system that is used in the overall system for monitoring the approach briefing. To test the robustness of the speech recognition part, a reasonable approach would be to evaluate the performance of a sufficiently trained language model in real cockpit environments. Their accessibility, however, is very limited in the scope of this project. In order to still be able to perform extensive testing under realistic conditions, several minutes of real cockpit noise were recorded using a high-grade digital audio recorder. The noise was captured on the flight deck of an Airbus A320 during cruise flight at the position where the pilot's headset microphone is usually located.

The cockpit noise was later mixed with human speech that was recorded from participants in a noise-free cockpit simulator. Due to the realistic size and geometry of the simulator, it was assumed that the acoustic properties are very similar to those of the real cockpit. In preliminary trials it was observed that participants in the cockpit simulator spoke differently when there was no noise present. Therefore, the participants were provided with real cockpit noise via their headsets, while the recording of their speech remained almost noise-free.

Having separate recordings of real cockpit noise and noise-free human speech allows for an easy generation of realistic data for training and testing the language model without the need to access real aircraft cockpits. By gradually deteriorating the signal-to-noise ratio, the robustness of the language model in noise intensive cockpit environments can be assessed. This process is ongoing and will serve as a quality measure for future iterations of the ASR system.

8. Towards a full Briefing Assistant

The recognition of speech in noisy cockpit environments with the help of an ASR system is the first step towards an advanced Approach Briefing Assistant. The assistant should be able to detect topics from spoken language either by directly recognizing specific keywords or by using the semantic context of words in order to deduct the content. This would allow the assistant to analyze an approach briefing even if the recognized transcript of the pilot's speech is not fully complete. Based on the detected topics the assistant can warn the pilots about possibly missing items they forgot to talk about.

Such a system will consist of three parts: the ASR (to recognize speech, as described in this paper), the topic model (to detect the topics addressed) and a check for completeness (to give pilots a feedback about the completeness of their briefing). Whereas a check for completeness can easily be done by comparing the recognized topics to a list of required entities, the main focus after the ASR is the topic model and the detection of the corresponding entities. The proceeding is to get the recognized but unstructured text of the approach briefing, extract the needed ontology vocabulary (i.e. which topics consist of which words) and create an entity list for them.[6)] The main difficulty here is the wide variance of words and phrases in the pilots' briefings when they talk about the same topics but use their own words. Applying AI-powered word embedding technologies (e.g. word2vec[7,8)]) one is able to automatically derive synonyms and alternative phrasing for the same context and later map the respective words to the same entities, hereby solving the problem described above.

Language Model	Perplexity	WER
ViCKI Trigram	**18.47** $\pm$ 2.79	17.88 $\pm$ 0.006
ATC Trigram	358.44 $\pm$ 27.92	54.60 $\pm$ 0.052
Brute Force Combination	31.88 $\pm$ 5.40	19.40 $\pm$ 0.011
Linear Interpolation	23.04 $\pm$ 3.36	**17.65 $\pm$ 0.007**

Table 1. Language model performance. The perplexities are reported on 11-fold cross Validation. WER is reported for three test briefings

The extension of the ASR to a full briefing assistant also has the potential to increase the accuracy of the speech recognition itself. By detecting the addressed topics and recognizing them as completed according to the reference briefing structure, it is possible to create a back-channel to the ASR system. Words that are usually contained in already completed sections of the briefing have a lower probability to appear again in the remaining parts of the briefing. This allows to lower the probability for recognizing these words by the ASR and increases the probability of other words to be recognized.

In the future, this work could lead to an even more sophisticated Approach Briefing Assistant that not only checks for completeness but also for correctness of specific briefing items. This could be accomplished by incorporating additional data sources like e.g. aircraft position, weather data or NOTAMs. Based on these data sources, the relevance of the various briefing items in a particular situation could also be derived, which would pave the way for dealing with newer and smarter flight crew briefing techniques (see section 10.).

9. Steps towards certification

Beside the technological aspects, another challenge is to certify such a system within the framework of aviation regulations. As established methods of system- and subsequent development could not be applied, the EASA is working with the industry and research institutions on guidance to define equivalent methods for the use of machine learning (ML). The fundamental EASA AI Trustworthiness Analysis, based on the EC Ethical Guidelines,[4)] is supported by three principles: Learning Assurance, AI Explainability and AI Safety Risk Mitigation.[2)] A first guideline[3)] has just been released, which addresses an initial application group of ML systems. The EASA distinguishes three major levels of ML systems: 1. Assistance to human; 2. Human/Machine collaboration; 3. More autonomous machine. The guideline uses common methods like configuration management and validation, as well as new methods addressing the specific characteristics of ML systems regarding the training stage. This has to be seen supplementary to the established development methods.

The approach briefing assistant has undergone an initial analysis according to the guideline and belongs to level one, assistance to human. During the analysis of the approach briefing assistant and the related training process new topics have been identified that are currently not covered by the EASA guideline. These topics need further discussion with the EASA. For example, the guideline assumes the full control over the training data sets. We consider the approach briefing assistant to be a good entry point into ML technology, as it does not interfere with the current procedures in the cockpit. The acceptance by pilots and the process of building trust in such ML based systems have to be further evaluated in the course of this project.

10. Future briefing techniques

With respect to the upcoming philosophy change in flight crew approach briefings (from content-based to threat-based),[1)] it is necessary to mention that the discussed speech recognition system is not capable of checking these new types of briefings against completeness without further adjustment. The new briefing concept focuses on threats and errors which are identified by each individual flight crew member. These threats are highly variable in terms of the individual experience of the crew member and the present approach situation (terrain, weather, fatigue, etc.). Therefore, from the design perspective, the term complete becomes more dynamic, which makes it in turn more difficult to check briefing items against completeness.

In general, the initiative to rethink the philosophy of conventional briefings was known to the authors, especially the new Airbus policy to implement Smarter Briefings. However, during the development process it was necessary to take the final decision which philosophy should form the basis for the speech recognition algorithm and accordingly the training data. When the design freeze was made, the current briefing structure used by major German airline carriers was the conventional briefing structure. Therefore, it was decided to follow the traditional content-based approach, since pilots still conducting conventional briefings represent the largest group of potential participants for training data generation.

The use case approach briefing can be understood as a demonstrator to show that flight crew speech recognition in noise polluted cockpit environments is possible. As stated before, this capability can be adapted for similar use cases, such as monitoring briefings according to the new philosophy.

11. Conclusion

This paper described an approach to implement AI-related functions within a cockpit environment. The use case of monitoring the approach briefing seems to enable a short-term demonstration of the functionality of such a system and also covers first considerations regarding the certification, which is very important in the aviation domain. Even with the small amount of interview-collected training data the results of the speech recognition model are very promising. The work on data collection is meant to be continued in order to further optimize the model.

Contributions

S. B. developed the concept for the incorporation of AI methods into the monitoring of an approach briefing. J. F. designed the briefing scenario and integrated the described system into the simulation environment of the Institute of Flight Guidance. C. R. brought up the idea to monitor the approach briefing through an AI-based system. Being a commercial pilot himself he collected most of the sample briefings from participating pilots. A. K. trained and continuously improved the ASR language model. N. G. helped in developing the concept of extending the ASR by using information extraction and word embedding to a briefing assistant. S. M.-D. led the discussions with the European Union Aviation Safety Agency (EASA) and helped to locate the system within the the framework of the EASA Artificial Intelligence roadmap. T. F. was one of the initiators of the ViCKI project and coordinated together with P.H. the research activities on the side of the Technical University of Braunschweig. S.O. organized the basic access to the ASR system by Cerence, contributed in discussions on the language modeling and co-managed the project on DFKI side. DK managed the development of the ASR component and contributed in discusssions.

Acknowledgments

The authors wish to thank the European Union Aviation Safety Agency for the fruitful dialogue on the steps towards a possible certification of the described system according to the EASA Artificial Intelligence Roadmap. Further, we would like to thank the company Cerence for providing a license for the VoCon ASR system for research purposes. All of us are very grateful to the many pilots who have contributed and will continue to contribute their approach briefings for the training and improvement of the language model.

This work was funded by the ViCKI project within the framework of the German Civil Aviation Research Programme (LuFo VI-1) of the Federal Ministry for Economic Affairs and Climate Action (BMWK) under grant number FKZ 20D1910D.

References

1) AIRBUS - WorldWide Instructor News: *What about Flight Crew Briefings?*, https://www.airbus-win.com/wp-content/uploads/2021/11/flight-crew-briefings-en-1.mp4

2) EASA: *Artificial Intelligence Roadmap, A human-centric approach to AI in aviation*, European Union Aviation Safety Agency, Cologne, 2020.

3) EASA: *EASA Concept Paper: First usable guidance for Level 1 machine learning applications*, European Union Aviation Safety Agency, Cologne, 2021.

4) EC: *Ethics Guidelines for Trustworthy AI. s.l.*, European Commission, 2019.

5) Hofbauer, K., Petrik, S. & Hering, H.: *The ATCOSIM corpus of non-prompted clean air traffic control speech*, Proceedings Of The Sixth International Conference On Language Resources And Evaluation (LREC'08), 2008.

6) Lodex: *Linked Open Data Extraction Tool enables to extract large ontology vocabularies out of unstructured data such as document text or social network messages*, https://git.opendfki.de/reuschling/lodex

7) Mikolov, T. et al.: *Efficient Estimation of Word Representations in Vector Space*, Proceedings of Workshop at ICLR, 2013.

8) Mikolov, T. et al.: *Distributed representations of words and phrases and their compositionality*, Advances in Neural Information Processing Systems, 2013.

9) Srinivasamurthy, A., Motlicek, P., Himawan, I., Szaszak, G., Oualil, Y. & Helmke, H.: *Semi-supervised learning with semantic knowledge extraction for improved speech recognition in air traffic control*, 10.21437/Interspeech.2017-1446.

10) Stolcke, A.: *SRILM-an extensible language modeling toolkit*, Seventh International Conference On Spoken Language Processing, 2002.